Online Appendix:
"Rushed Innovation: Evidence from Drug Licensing"

By Manuel Hermosilla

October, 2019

A. Computation causal forest CATE estimates

The CATE estimate for observation i, τ_i, is obtained by fitting the following weighted moment condition:

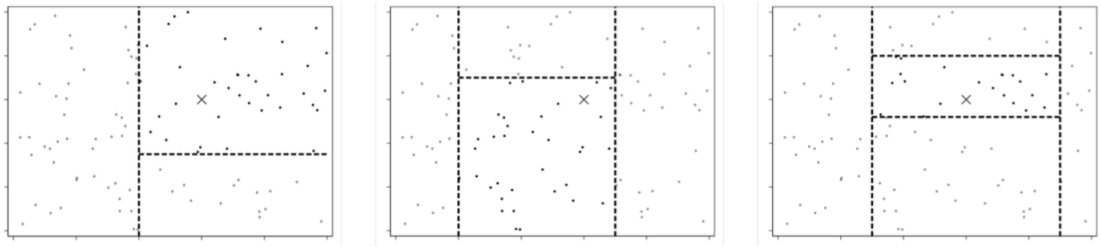
$$E\Big[(Y_{i'} - m_{i'}) - (W_{i'} - \pi_{i'}) \cdot \tau_i\Big] = 0,$$

where $m_{i'}$ corresponds to a random forests prediction of Y_i. The expectation is taken over all observations i' in the sample, each weighted according to its similarity score $\alpha_i(i') \in [0, 1]$. Thus, τ_i is estimated from observed deviations of outcomes and treatment assignments (from their respective predictions), using a dataset that is specific to observation i (i.e., observations weighted according to their similarity to i). Also computed through random forests, these weights identify similar observations based on variables' actual impacts on the outcome of interest (covariates which do not explain the outcome do not decrease similarity between observations), while flexibly accounting for non-linearities and discontinuities in these relationships.

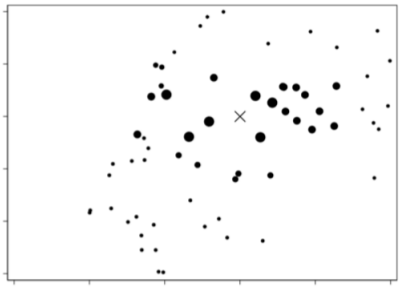
Although random forests are used to generate the predictions m and π, their essential role in generalized causal forests is the computation of similarity weights. A large number of classification trees (2,000 in our case) separately characterize the relationship $Y = f(X)$, where X corresponds to the set of variables to which CATE heterogeneity is pegged. These trees are random, in the sense that they base classification on random subsamples and subsets of X. Each tree produces a partition of the data, such that each resulting subset is characterized by a relatively homogeneous outcome Y. Panel A of Figure A.1 illustrates what these partitions would look like when observations are classified based on the values of two covariates, X^1 and X^2 (vertical and horizontal axis), for three hypothetical trees. The first tree (left) classifies results in a 3-fold partition (each subset is called a "leaf"). The observation marker with an "$\times$" symbol will then be associated to the same predicted value of Y as all other observations in the same leaf leaf (top right). Tree-level similarity is formulated dichotomously: $\times$ is equally similar to all observations in the same leaf, and non-similar to all other observations. Random trees on the center and right positions of the panel produce different partitions, and therefore associate each observation with different sets of similar observations. The similarity weight of observation i' with respect to observation i, $\alpha_i(i')$, is computed as the percentage of trees in which i' is classified into the same leaf as i. The Figure's Panel B shows the resulting set $\{\alpha_\times(i')\}$, where larger markers indicate weights of higher magnitude. Weights for observations in the left and bottom right are zero because our three hypothetical random trees do not classify these observations into $\times$'s leafs.

Figure A.1: Similarity weights used for the causal forest estimator: illustration for observation $\times$.

A. Classification in three random trees



B. Similarity weights (weight magnitude given by marker size)

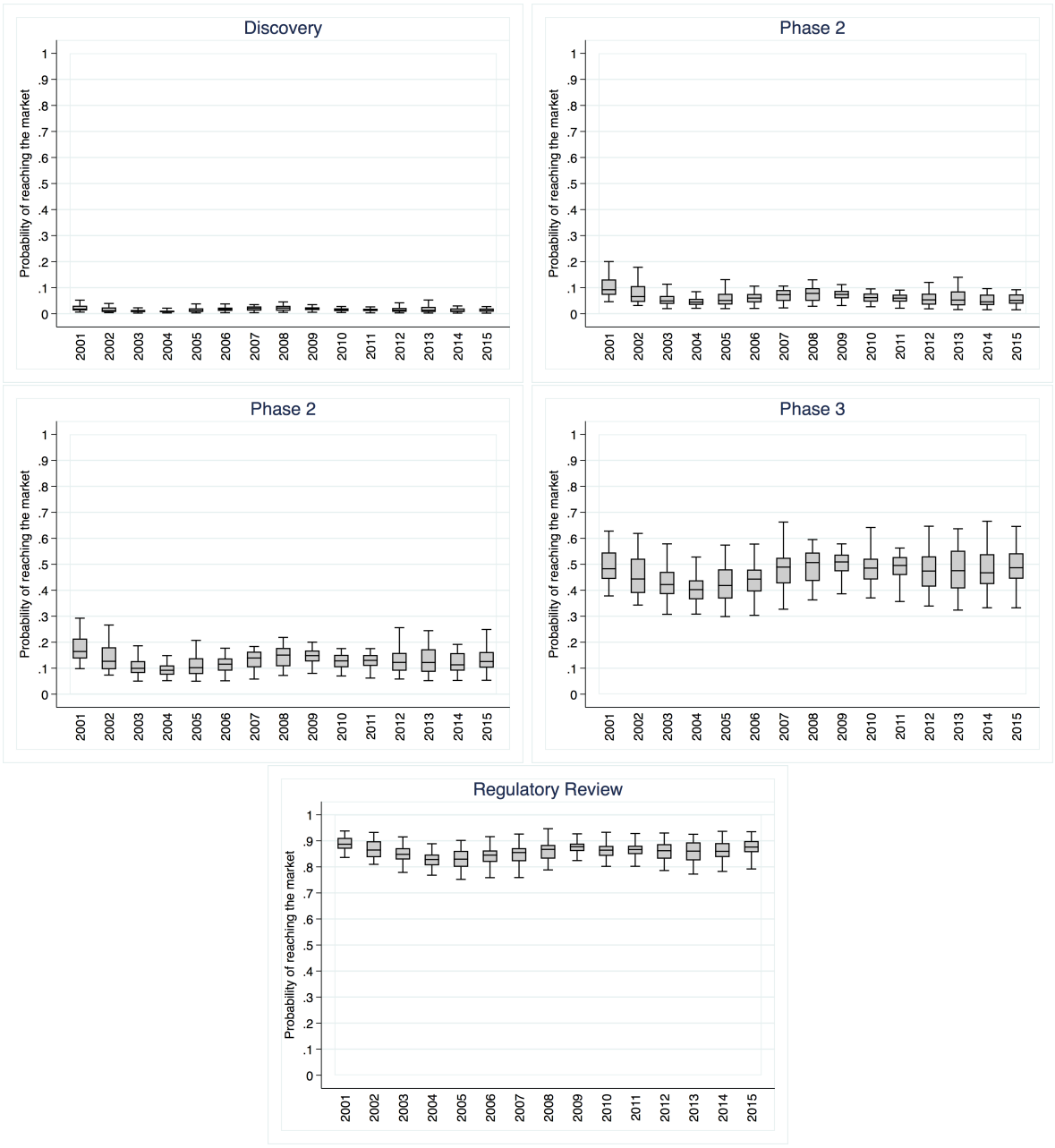


B. (Stage-conditional) Reach-the-market probabilities

Here we describe the procedures used to construct the set of "(stage-conditional) reach-the-market" probabilities $\{p_{as}\}$ used for the weighted aggregation described in **??**. A probability p_{as} corresponds to the probability that a therapy of area a and which is at stage s reaches the market some time in the future (as opposed to having it development terminated sometime along the way). In addition to differencing across areas and stages, the procedure used differences across time. That is, estimated probabilities are in fact associated to a third subindex y referring to the year associated to the counts that are aggregated (e.g., if we are measuring pipeline strength in some quarter of 2001, probability weights $\{p_{as2001}\}$ are used). We construct these probabilities on a rolling basis and at the therapeutic area level, using the about 80,000 development histories available from the broad Cortellis sample. Importantly, none of the development histories analyzed in the paper are used to construct these probabilities.

To construct the probabilities used to measure counts associated to a date t, we will only consider subset of the outcomes entailed by the over 80,000 development histories in the data. In particular, we consider outcomes were observed between y-5 and y-1, with y =year(t). With these, we estimate a probit model for the probability of advancement for one stage to the next (as opposed to termination), considering one stage at the time, and using therapeutic area indicators as independent variables. The predictions generated by these estimates are stored, and the estimation is repeated for all other stages and years in the sample. Call the resulting probabilities q_{asy}. Stage-conditional reaching-the-market probabilities are then computed as $p_{asy} = \prod_{s' \geq s} q_{as'y}$. Figure A.2 describes these estimates, for each year in the sample (boxes reflect variability across therapeutic areas). Estimated probabilities are quite stable across the sample period.

Figure A.2: Estimated (stage-conditional) reach-the-market probabilities.



C. Balancing statistics for propensity score matching

To assess the balancing of matched samples we use standardized differences, which Rosenbaum and Rubin (1985) first used to evaluate how comparable treated and control observations were in a matched sample. The approach has then become popular in the causal inference literature, one reason being that, compared to t-tests and other statistics that evaluate the significance of group differences, standardized differences have the advantage that they are not influenced by sample size (Austin, 2009). Another advantage is that the approach deals naturally with the comparison of variables cast in different scales.

The standardized difference for a variable X_k is defined as:

$$d_k = 100 \cdot \frac{\bar{X}_k^T - \bar{X}_k^C}{\sqrt{\frac{\sigma_k^T + \sigma_k^C}{2}}},$$

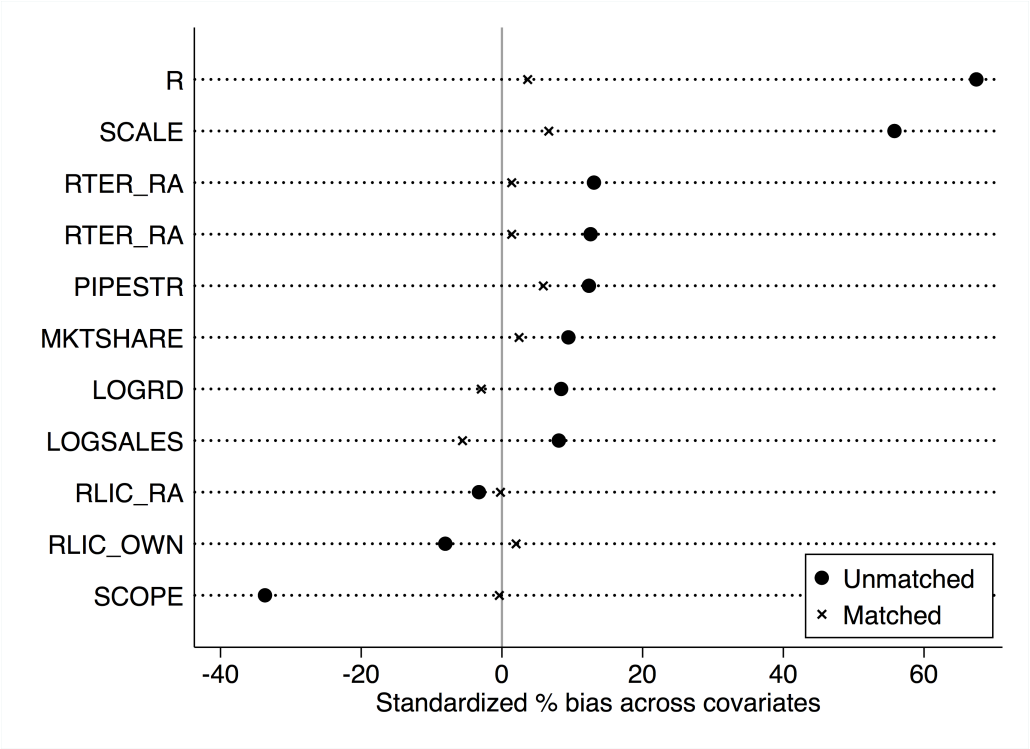
where T indicates that an observation belongs to the treatment group and C that it belongs to the control group. $\bar{X}$ corresponds to averages and σ to sample variances.

One challenge of standardized differences is that there exists no wide consensus on the thresholds that would denote substantial treated/control covariate imbalances (Austin, 2009). We will consider the prescription of Cohen (1988), which indicates that standardized differences thresholds for small, medium, and large imbalances can be implemented respectively by 20, 50, and 80.

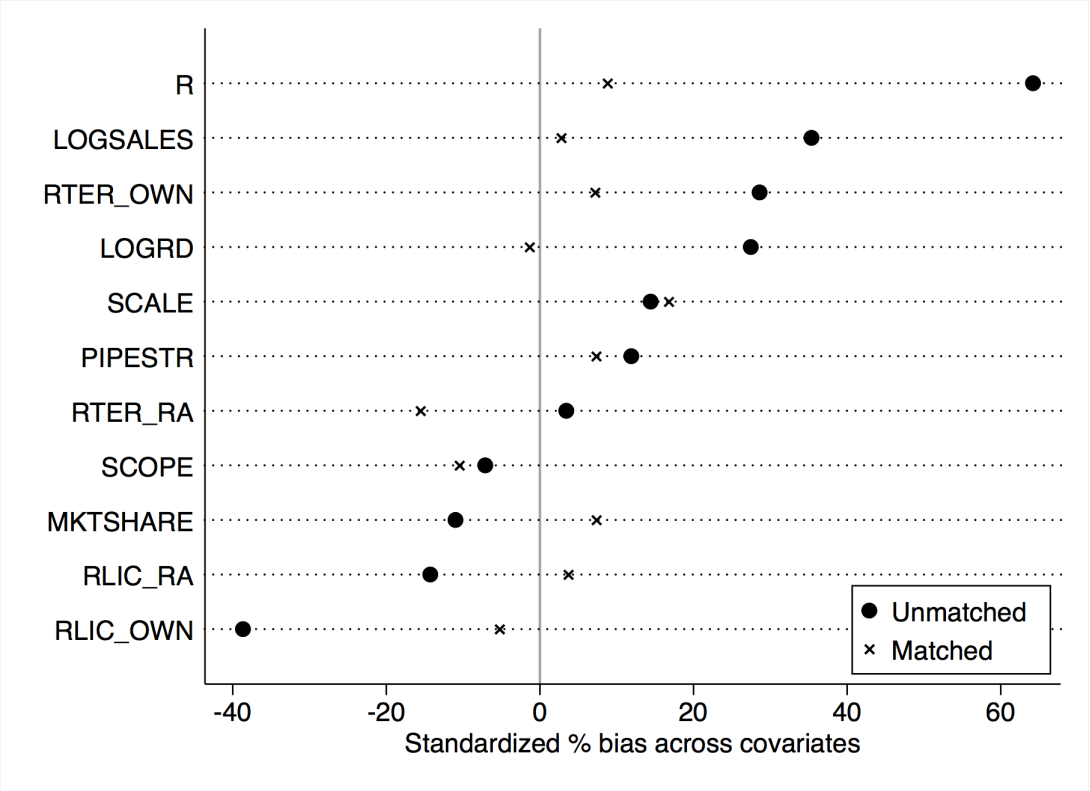
Figure A.3 presents the balancing results. Particularly in the sample used for the performance analysis, some variables display large imbalances in the unmatched sample. These are sharply reduced in the matched samples, within which all variables display standardized differences below the "small imbalance" threshold of 20.

Figure A.3: Balancing statistics for propensity score matching models.

A. Licensing analysis



B. Performance analysis



D. Probing the influence of non-standard P3F shocks

In **??** we described the anatomy of P3F events, highlighting a series of nuances that are responsible for "non-standard" P3F shocks. The analyses presented in the paper's main text assumed these nuances away. Here we revisit these issues and probe their influence on our inference.

We begin by re-stating the issues raised in **??**. "Non-standard" P3F shocks can arise if they:

S1. Are composed of the Phase 3 failure of more than one therapy.

S2. Are composed of the Phase 3 failure of therapies associated to more than one compound. (This issue was not made explicit in **??** but is a direct consequences of the premises therein.)

Both these scenarios are a logical consequence of aggregation that is necessary to implement our research design. The next two scenarios are consequences of the fact that many compounds are developed to treat more than one condition (i.e., a single molecular entity can have more than one therapy tested in Phase 3). In these cases, P3F shocks will be "non-standard" if:

S3. A compound's multiple therapies that "belong" to different therapeutic areas fail within the same measurement window, implying that more than one decision-making unit will be impacted.

S4. Therapies drawn from the same compound register failures over different measurement windows. This would mean that latter failures represent P3F shocks that have been "preceded" by another P3F shock from the same compound.

These scenarios imply affected P3F shocks are not independent. Note that scenarios S1-4 are not mutually exclusive by construction.

Table A.1 presents the results of our sensitivity analysis for each of these scenarios (Columns 2-5). Let us first focus on the row "% treated obs dropped," which describes the percentage of treated observations that are affected by each. Panel A focuses on the data formatted for the licensing analysis; Panel B, on that for the performance analysis. Recall that the measurement window is one quarter for the licensing analysis. For the performance analysis it is one year.

The statistic of Column 2 (Panel A) indicates that 16% the P3F shocks codified for the licensing analysis were composed by the Phase 3 failure of more than one therapy. The statistic of Column 3, that in over a third of these (6% overall) failed therapies were associated to different compounds. The statistic of Column 4 indicates that 11% of P3F shocks represent cases in which two or more therapies of a same compound but which "belong" to different areas fail during the same window. Lastly, the statistic of Column 5 tells us that 10% of P3F shocks are "repeated," meaning that, for them, a previous therapy of the same compound had failed in Phase 3 in a previous quarter. The more frequent scenario is S1 (multiple therapies). Panel B presents the same statistics but computed for the performance analysis. The aggregation window is now four times as wide (a year), hence percentages are significantly larger. Except for scenario S1 (which remains the more frequent), the percentages of treated observations that are affected remain below 20%.

To investigate how much the above issues influence our main results we first reproduce our linear probability estimates (Column 2 of **??** and **??**) but successively dropping the treated observations that are affected by each issue. Recall that this is an specification in which we regress the outcome (DLICENSE or ADVANCE) on a P3F treatment indicator (reported estimate) and a wide array of contextual factors and fixed effects. Overall, results of Columns 2-5 are similar to the estimate of our main analysis (this baseline estimate is reproduced in Column 1 for convenience).[1] In fact, effects are generally reinforced. The exception is that estimate of S1 in the licensing analysis (Column 2, Panel A). In this case, the coefficient's statistical significance level falls marginally below the threshold of 90% confidence. Because issues are not mutually exclusive, as an additional check we enrich our specification with four indicators (one for each scenario), and estimate it on the full sample (no observations dropped). Again in this case, reported estimates (Column 6) continue to support our conclusions. We conclude that our main results are not driven by the confounds introduced by "non-standard" P3F treatments.

Table A.1: Probing the influence of non-standard P3F treatments.

	(1)	(2)	(3)	(4)	(5)	(6)
		Exclusion condition for treated observations				
	Baseline	Multiple therapies (S1)	Multiple compounds (S2)	Multiple areas (S3)	Repeated failure (S4)	Inidicators for each scenario added (S1-4)
A. Licensing analysis (outcome DLICENSE)						
Estimate	0.033*	0.034	0.038*	0.053**	0.052**	0.046**
Std. error	(0.020)	(0.022)	(0.021)	(0.021)	(0.021)	(0.023)
% treated obs dropped		16%	6%	11%	10%	
Obs.	6,479	6,412	6,455	6,432	6,436	6,479
B. Performance analysis (outcome ADVANCE)						
Estimate	-0.036*	-0.094**	-0.064*	-0.062**	-0.054**	-0.107**
Std. error	(0.020)	(0.035)	(0.030)	(0.026)	(0.024)	(0.040)
% treated obs dropped		42%	17%	11%	16%	
Obs.	1,783	1,664	1,733	1,752	1,737	1,783

Evidence from linear probability models. The specification of Panel A corresponds to that of Column 2 in **??**. The specification of Panel A corresponds to that of Column 2 in **??**. Robust standard errors are presented in parentheses. Legend: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

[1] In the main text, this baseline estimates are presented in**??** and **??**, Panel A, Column 2.

E. Do licensing events impact stock market returns of the in-licensing firm?

To address this question, we employ an event-study methodology that has been applied numerous times to study how innovation-related events impact firms' stock market returns.[2] Our results suggest that (i) licensing events do not impact the in-licensing firm's stock market valuation, and (ii) pre-licensing P3Fs do not appear to systematically mediate these impacts.

The analysis relies on an "abnormal returns" variable (AR), which is computed at the daily level, based for the in-licensing firm's stock price returns. Inference is based on the accumulation of these abnormal returns around an "event day," which in our case corresponds to the day in a therapy j is in-licensed. We index this day by $t = 0$. Days following the event are successively indexed by $t = 1, 2, ..$; those that precede it, by $t = -1, -2, ..$. AR represents the difference between the stock's actual daily return and that predicted by an equilibrium pricing model.

Crucially, the event-study methodology postulates that ARs account for the impact of the information surfaced by the event under study on the firm's stock market returns. Formally, inference is drawn by aggregating ARs into cumulative abnormal returns (CAR) over the event window $\{\underline{T}, .., \bar{T}\}$. To analyze the impact of a therapy j's licensing on its in-licensing firm's stock returns, CAR is defined as:

$$\text{CAR}_j = \sum_{t=\underline{T}}^{\bar{T}} \text{AR}_{ft},$$

where f indexes the in-licensing firm. Following other studies in this literature (e.g., Girotra et al., 2007), we set $\underline{T} = -1$ to account for the possibility that there may have been "information leakages" prior to each deal's announcement. Keeping this threshold fixed, we consider the three windows, given by $\bar{T} = 3, 5, 7$. Thus, for all considered windows, CAR computes the sum of abnormal returns experienced by the firm around the deal's date. An estimated value of CAR > 0 would be consistent with the idea that licensing a positive impact on the firm's stock market valuation.

To compute ARs, we consider two equilibrium pricing models to predict expected returns $\mathbb{E}[\text{R}_{ft}]$. With these predictions, ARs are constructed as AR_{ft}=R_{ft}-$\mathbb{E}[\text{R}_{ft}]$. The employed pricing models are the Market Model (MM) of Brown and Warner (1985) and the Fama-French-Cahart 4-Factor Model (FFC) of Carhart (1997). These compute expected returns respectively as:

$$\begin{aligned} \mathbb{E}[R_{ft}] &= \alpha_j + \beta_j R_{mt} && \text{(MM)} \\ \mathbb{E}[R_{ft}] &= \alpha_j + \beta_j R_{mt} + \theta_j^1 \text{SMB}_t + \theta_j^2 \text{HML}_t + \theta_j^3 \text{UMD}_t && \text{(FFC)} \end{aligned}$$

Both models produce expected return predictions by invoking an equilibrium condition that ties the return of the market portfolio (R_{mt}) to that of the firm's stock. The parameters governing this relationship are estimated independently for each licensing event

[2]For example, Sharma and Lacey (2004) use them to unveil the asymmetry of reactions to good and bad news in the pharmaceutical industry, while Girotra et al. (2007) to dissect the negative impacts of P3Fs based on the richness of the firm's developmental pipeline. Chaney et al. (1991) provide estimates of stock market reaction to developmental success across industries. Cao and Sorescu (2013) focus on the impact of co-branded products. Hendricks and Singhal (1997) study the impacts of product delays, and Robertson et al. (1995) how new product announcements impact the stock market valuation of competing firms. See Srinivasan and Hanssens (2009) for a thorough review of these methods and their application to marketing problems

through OLS regressions, using data from the pre-licensing window $\{-365, .., -11\}$. Models differ in terms of the number of additional factors that enter the specification as independent variables. Whereas MM only includes return of the market portfolio, FFC includes three other potential moderators. These are, (i) SMB ("small minus big") factor (return spread between small and large firms), (ii) HML ("high minus low" or "value premium") factor (differential return between stocks of companies with high and low book-to-market ratios), and (iii) UMD ("up minus down") factor (historical excess return of "winning stocks," compared to "loosing" ones). Factor data was obtained from Professor Kenneth French's website; stock market returns data, from the Center for Research in Security Prices (CRSP). We present results for the 1,775 licensed therapies for which we were able to compile all required data.

Estimated CAR values are small on average, suggesting that licensing does not have a systematic impact on the in-licensing firm's stock market value. When MM is used, these average 0.001, 0.002, and 0.004 for each window (narrowest to widest). Their FFC counterparts are about an order of magnitude smaller. Between 2% and 6% of these estimates are statistically significant with 95% confidence. (Standard errors are obtained from each regression's estimated error variance). These relatively few statistically significant CAR estimates are about evenly split between positive and negative values.

To further illustrate the absence of a systematically positive effect of licensing events on the in-licensing firm's stock market valuation, Figure A.4 presents the (cumulative) distribution of estimates obtained in each case. To investigate the possibility that these effects may be present when licensing follows a recent P3F, we present separate distributions for each set of events (i.e., with and without a P3F having occurred within year ending on the licensing date). If the referenced impacts of licensing existed, these distributions should place relatively more mass on the positive domain of CAR. Plots show that this is not the case. Moreover, the incidence of a pre-licensing P3F does not appear to make a difference: dashed (P3F) and solid (no P3F) distributions are quite similar. The hypothesis that CAR estimates for licensing events that follows a P3F are larger than their "no-P3F" counterparts is formally rejected within all sets of results (95% confidence).

As a last check, we evaluate the volume of stocks traded around the date of licensing $t = 0$. We base our analysis on the variable NVOL, defined as:

$$\text{NVOL}_{jt} = \frac{\text{VOL}_{jt}}{\bar{\text{VOL}}_j} - 1 \qquad \text{for } t = -10, .., 0, .., 10,$$

where VOL_{jt} represents the number of stocks traded on day t (of therapy j's in-licensing firm). The denominator $\bar{\text{VOL}}_j$ corresponds to the the average of VOL within the period $t = -365, .., -11$. Thus, NVOL gives a normalized measure of traded volume around licensing. If licensing has an impact on the in-licensing firm's equity, we should expect NVOL values to shift around licensing. The patterns described by Figure A.5 suggest that this is not the case. NVOL values are quite stable around zero throughout the considered window. In conclusion, this evidence does not support the idea that licensing impacts the in-licensing firm stock market valuation.

Figure A.4: Distributions of estimated CARs around in-licensing events.

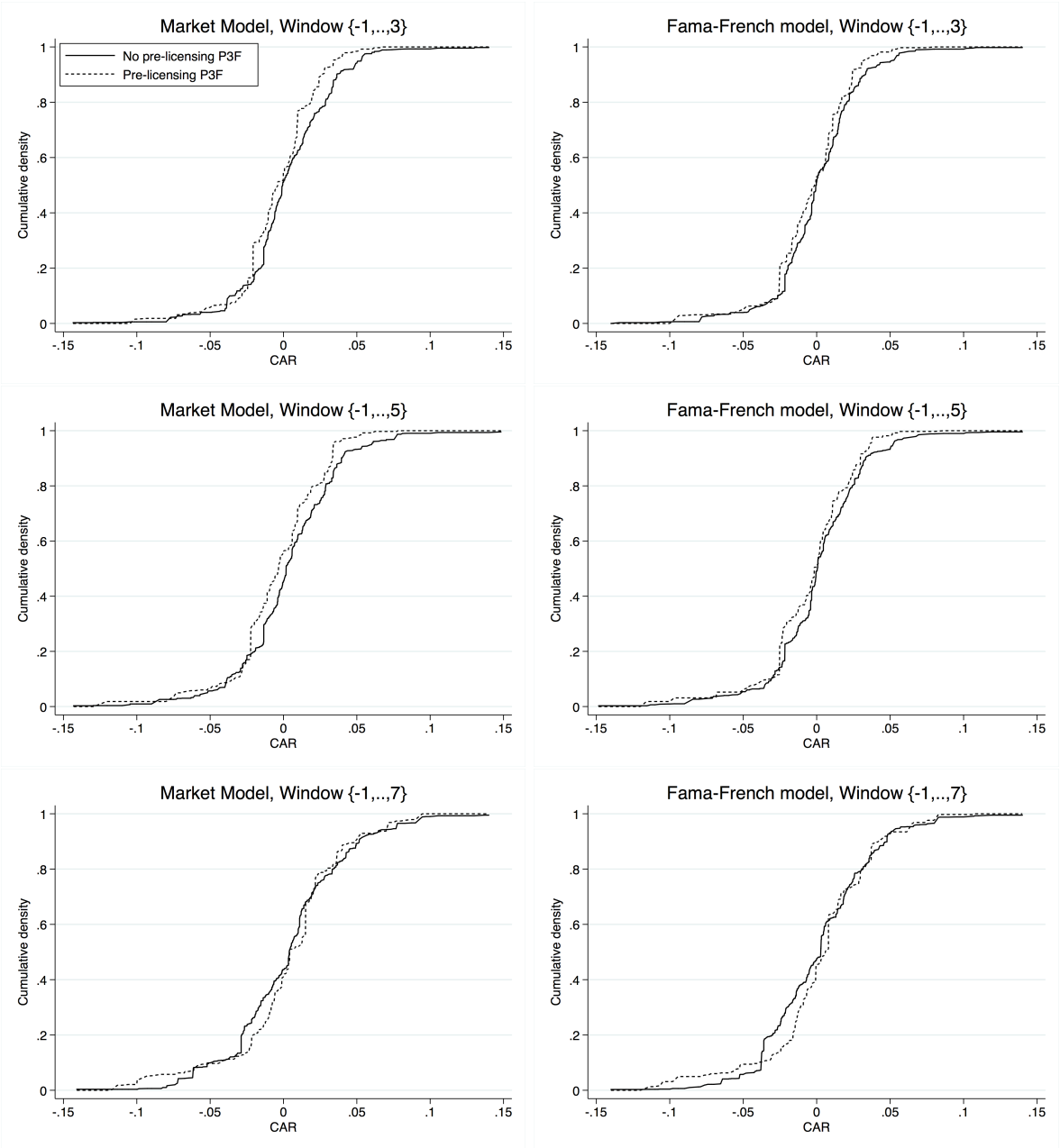
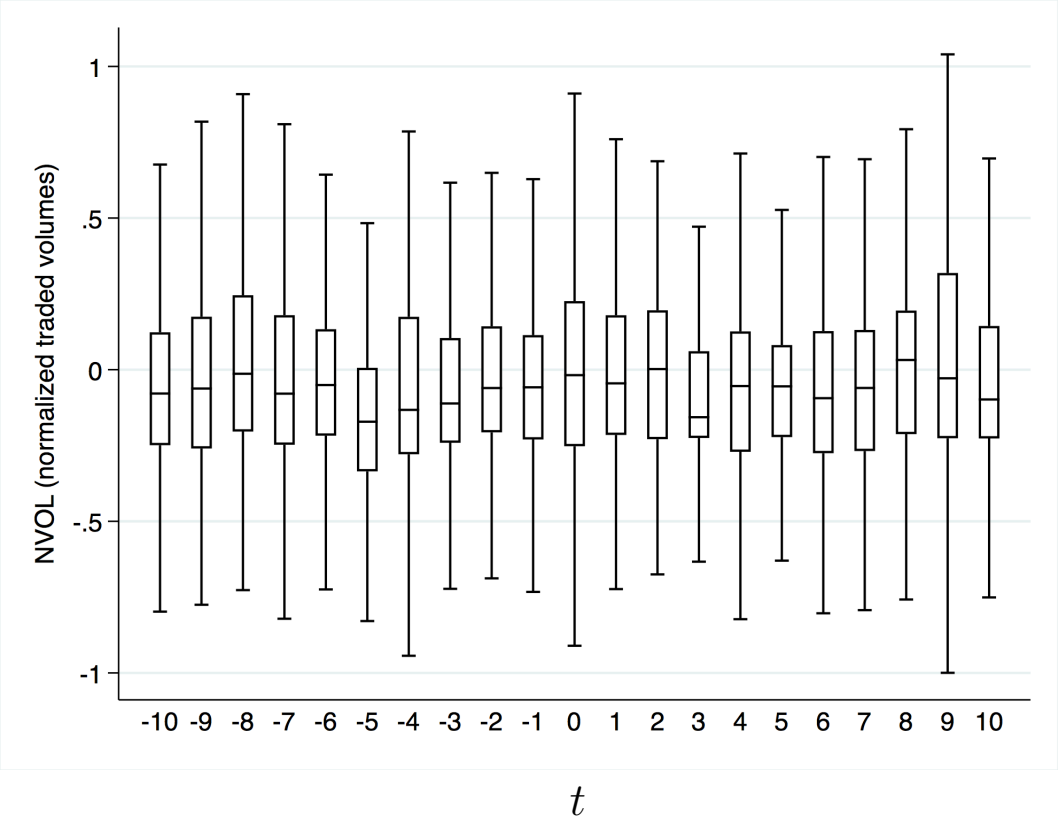


Figure A.5: Normalized traded volumes (NVOL) around the licensing date ($t = 0$).



F. Investigating "capture and kill" behavior

One challenge to our interpretation of the performance results of **??** stems from potential "capture and kill" behavior. Under this rationale, rushed licensing would not necessarily aim at filling a pipeline gap, but instead, at hindering future competition by "killing" therapies that are available for competitors to license.[3] An important difficulty to investigate this claim is that the analysis cannot rely on development outcomes—the two hypotheses are observationally equivalent in this dimension. We thus rely on a rationale based on the timing of milestone completion. Our reasoning is: since firms could achieve the "capture and kill" objective by immediately shelving licensed therapies (further investment in costly clinical trials is not justified), development outcomes of therapies that are "captured and killed" should be observed earlier (relative to the date of licensing). We test this implication with the following specification:

$$\text{CMONTHS}_k = \beta_0 + \beta_1 \cdot \text{ADVANCE}_k + \beta_2 \cdot W_k + \beta_3 \cdot (1 - \text{ADVANCE}_k) \cdot W_k + \Theta X_k + \epsilon_k,$$

where k indexes licensed therapies and CMONTHS corresponds to the number of months that separate licensing from the completion of the first post-licensing milestone. (Our analysis will only consider each therapy's first post-licensing outcome.) As such, smaller values of CMONTHS indicate that post-licensing outcomes are observed earlier. W is the treatment indicator (i.e., pre-licensing P3F), ADVANCE is defined as before, and X contains all the controls and fixed effects included in the linear probability models of **??**. The coefficient of interest is that attached to the interaction, β_3. If the underperformance of treated therapies was driven by "capture and kill" behavior, we should observe that termination outcomes for treated therapies (i.e., "killings") are observed earlier. That is, we should obtain a negative estimate for β_3. In contrast, the estimated parameter takes on an imprecisely estimated positive value ($\hat{\beta}_3 = 6.170$, $p = 0.204$). Thus, this result suggests that the therapies licensed in a rush are not terminated any sooner than those licensed under normal conditions (no pre-licensing P3F). We therefore conclude that "capture and kill" behavior is unlikely to drive our performance results.

[3] A similar rationale has been recently put forth by Cunningham et al. (2018) in the context of pharmaceutical M&A activity.

G. Drivers of treatment effect heterogeneity

Here we seek to shed some light on the drivers of the treatment effect heterogeneity documented above. Our analysis is based on the associations of contextual factors with estimated treatment effects. To simplify our language, we refer to each of these associations as an "effect" (no causality implied). To set up our analysis it is first important to note that, since CATE heterogeneity is non-parametric, a contextual factor's effect may depend on those of other factors. For this reason we base our analysis on residual variation. More concretely, the "effect" of a contextual factor X^k (e.g., PIPESTR) on each CATE estimate $\hat{\tau}$ is assessed through the coefficient of correlation that arises between $\tilde{\tau}$ and $\tilde{X}^k$, where $\tilde{X}^k$ corresponds to the residual term obtained from a OLS regression of X^k on all other contextual factors X^{-k} (in addition to year and area fixed effects). In turn, the residual $\tilde{\tau}$ is obtained from the same specification but using $\hat{\tau}$ as dependent variable. As such, the "effect" of a contextual factor X^k is investigated through variability that is not explained by the remainder of the context. The process is repeated for CATE estimates from the licensing and performance analyses (labeled τ^L and τ^P, respectively). To compute residuals for the latter, we also include development and licensing stage fixed effects.

To facilitate our analysis it is useful to interpret CATE estimates as units' context-dependent "inclinations." Licensing estimates τ^L can be seen as units' inclinations to engage in rushed licensing. Larger (i.e., more positive) τ^L estimates point to P3Fs impacts of higher intensity. Similarly, τ^P estimates can be interpreted as units' inclinations to terminate the development of a therapy that was licensed in a rush. For these, smaller (i.e., more negative) values indicate higher-intensity impacts. As a further simplification we reverse the sign of τ^P estimates before computing residuals. In this way, larger (i.e., more positive) values of $\tilde{\tau}^P$ reflect impacts of higher intensities.

Panel A of Figure A.6 shows the resulting coefficients of correlation, or "effects." Contextual factors with larger licensing effects (black markers) tend to also be associated with larger performance effects (hollow markers). This pattern suggests that units' two types of inclinations may share a common structure. This observation is further analyzed below. Rest-of-area licensing (RLIC_RA) is the most notorious departure from this pattern. The negative performance effect (bottom left) indicates that a specific unit is less inclined to engage in rushed licensing when other units in its area have been in-licensing more actively. Given the limited licensing supply, this result could be explained by possible crowd-out effects. In turn, this crowd-out rationale would suggest that in circumstances like these, a unit would be less inclined to terminate a therapy that was licensed in a rush. However, we observe the opposite (bottom right marker). This result could be rationalized under the premise that rest-of-area licensing proxies for technological renewal.

Turning to portfolio-related variables, we first note that strong pipelines may be the reflection of a unit's strategy to launch a relatively large number of new therapies in the medium term. The positive licensing effect for PIPESTR could thus follow from the fact that given their strategic goals, units with stronger pipelines are more compelled to engage in rushed licensing. The effect could also follow from a more widespread availability of resources to these units. Since stronger pipelines also imply a higher degree of internal competition among developing therapies, units with stronger pipelines may also exercise more stringent thresholds when deciding on a project's termination. If so, units with relatively strong pipelines may be more inclined to terminate therapies licensed in a rush, as suggested by the strong performance effect for PIPESTR. The sign of these two effects is reversed for MKTSHARE. For units that enjoy a more robust market position, there

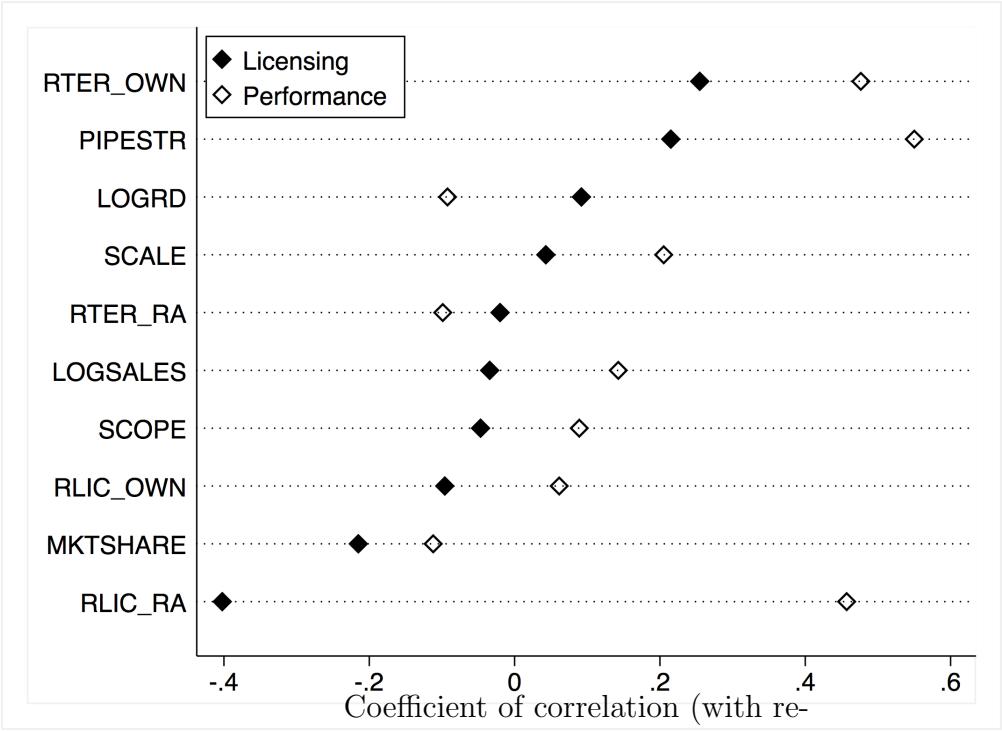
may be less pressure to supplement their pipelines following a P3F.

The effects for variables related to recent productivity also lend themselves to some interpretation. Poorer rest-of-area productivity (RTER_RA) may indicate the opening of a market opportunity (reduced number of future launches by competitors). As suggested by the performance effect of RTER_RA, units may be less inclined to terminate therapies licensed in a rush in these circumstances. Consistent with our previous analysis of "desperate" units, a unit's own poor productivity may increase its inclination to engage with rushed licensing (licensing effect for RTER_OWN—top right). The large and positive performance effect for own productivity suggests that units that experience poor recent productivity may be more inclined to terminate therapies licensed in a rush. This result is somewhat puzzling, however, as one would expect units in these circumstances to lean towards "salvaging" active projects. One possible explanation is based on strategic shifts. For example, consider a firm that decides to "retreat" from a certain area, like Pfizer did in 2008 by terminating several early-stage obesity and cardiovascular therapies (Kimes, 2008). This decision will directly worsen the firm's productivity record. If therapies licensed in a rush have relatively lower values of continued development, the retreat decision may also translate into a stronger inclination to terminate them.

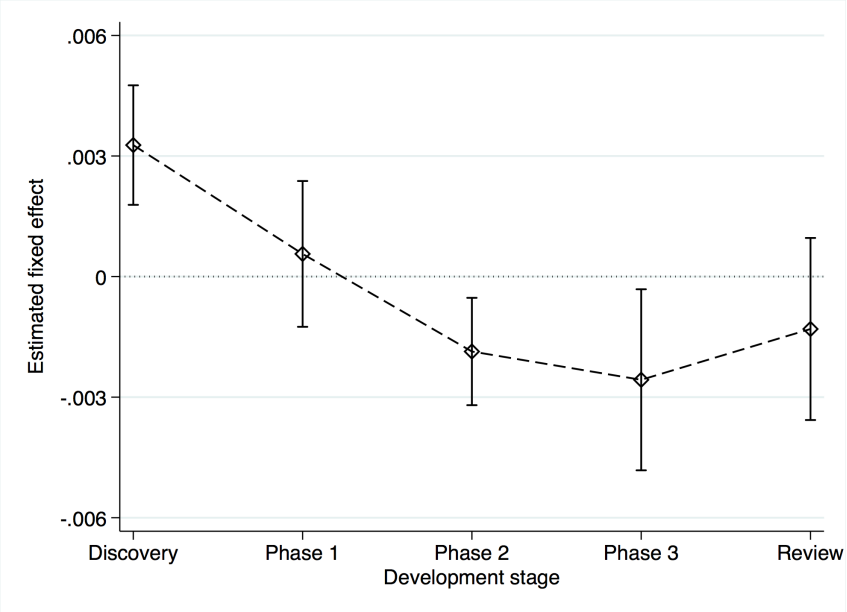
We conclude by comparing performance treatment effects across development stages. The analysis is motivated by the idea that the value of continued development is larger for therapies being developed at more advanced stages as these are (statistically) closer to the market. Units should therefore be less inclined to terminate a therapy licensed in a rush the more advanced its development stage is. We implement the analysis by re-estimating performance treatment effect residuals $\tilde{\tau}^P$ without including stage fixed effects so that the obtained residuals contain stage-specific effects. Panel B of Figure A.6 presents the estimates that we obtain when we regress these residuals on stage indicators. The (mostly) downward-sloping pattern indicates that the intensity of performance treatment effects decreases at more advance stages. That is, consistent with our intuition, units are less likely to terminate a therapy licensed in a rush when it is closer to the market.

Figure A.6: Drivers of treatment effect heterogeneity.

A. Correlating CATEs to contextual factors through residual variation



B. Performance CATE variation ($\tilde{\tau}^P$) across development stages.



Panel A. Markers reflect the intensity of the association between each variable and licensing (black) and performance causal forest CATEs (hollow). These are computed as coefficients of correlation between OLS residuals (details in text). **Panel B.** Computed as stage fixed effects in a OLS regression for residual performance cate variation (details in text). 95% confidence intervals shown.

References

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine 28*(25), 3083–3107.

Brown, S. J. and J. B. Warner (1985). Using daily stock returns: The case of event studies. *Journal of financial economics 14*(1), 3–31.

Cao, Z. and A. Sorescu (2013). Wedded bliss or tainted love? stock market reactions to the introduction of cobranded products. *Marketing Science 32*(6), 939–959.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance 52*(1), 57–82.

Chaney, P. K., T. M. Devinney, and R. S. Winer (1991). The impact of new product introductions on the market value of firms. *Journal of Business*, 573–610.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd ed.* Lawrence Erlbaum Associates, Publishers. Hillsdale, NJ.

Cunningham, C., F. Ederer, and S. Ma (2018). Killer acquisitions. *Available at SSRN 3241707*.

Girotra, K., C. Terwiesch, and K. T. Ulrich (2007). Valuing r&d projects in a portfolio: Evidence from the pharmaceutical industry. *Management Science 53*(9), 1452–1466.

Hendricks, K. B. and V. R. Singhal (1997). Delays in new product introductions and the market value of the firm: The consequences of being late to the market. *Management Science 43*(4), 422–436.

Kimes, M. (December 1, 2008). The case of pfizer. Fortune Magazine.

Robertson, T. S., J. Eliashberg, and T. Rymon (1995). New product announcement signals and incumbent reactions. *The Journal of Marketing*, 1–15.

Rosenbaum, P. R. and D. B. Rubin (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician 39*(1), 33–38.

Sharma, A. and N. Lacey (2004). Linking product development outcomes to market valuation of the firm: The case of the us pharmaceutical industry. *Journal of Product Innovation Management 21*(5), 297–308.

Srinivasan, S. and D. M. Hanssens (2009). Marketing and firm value: Metrics, methods, findings, and future directions. *Journal of Marketing research 46*(3), 293–312.